DOCUMENT RESUME

ED 246 101 TM 840 376

AUTHOR Clarke, Marian
TITLE Functional Level Testing Decision Points and Suggestions to Innovators.
PUB DATE Nov 83
NOTE 11p.; Paper presented at the meeting of the California Educational Research Association (Los Angeles, CA, November 17-18, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; *Decision Making; Educational Innovation; Elementary Secondary Education; Needs Assessment; *Program Development; *Testing Programs; Test Selection
IDENTIFIERS *Out of Level Testing

ABSTRACT

The innovators of a functional level testing program must make decisions in the context of situational factors. There are several decision points involving the important elements of a testing program. Establishing the need for functional level testing is the first decision to be made. A systematic examination of in-level test data will document the exact nature and scope of testing problems and form the foundation for future decisions. The innovators of functional level testing need to generate support for the testing proposal from persons empowered to prevent or facilitate its implementation. The most complex decision is selection of appropriate test levels for students. Test administration procedures may be questioned. Data suggest the feasibility of administering several test levels in one classroom. Scoring and processing procedures should be determined by cost, number of school personnel needed and turn-around time. On-site coordination is essential for the operational success of a functional level testing proposal. A systematic review of test score reliability must be a continuous practice. A feedback loop between the school sites and central office should be developed. References are listed in Appendix A and a series of achievement tests and recommendations for out-of-level testing are included in Appendix B. (DWH)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED246101

# Functional Level Testing Decision Points and Suggestions to Innovators

Symposium

Out-of-Level Testing: The Operational Reality

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. Clarke

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Marian Clarke, Consultant
Planning, Evaluation and Research
California State Department of Education

Presented at CERA

November, 1983

TM 840 376

FUNCTIONAL LEVEL TESTING DECISION POINTS

AND SUGGESTIONS TO INNOVATORS

Several decisions face the innovators of a functional level testing program. These decisions, made in the context of situational factors, will shape a unique testing program for the school system. Seven decision points that capture the important elements are: establishing need, generating support, selecting test levels, administering tests, scoring and processing, on-site coordination, and evaluation.

Decision Point #1
Establishing Need

The first and most important decision is to determine if functional level testing is needed. Teacher complaints that tests are too difficult or too easy may signal need, but a systematic examination of in-level test data will document the exact nature and scope of in-level testing problems and form the foundation for future decisions. Recommendations for conducting a systematic examination of existing data follow.

1.1 Establish criteria for reliable test score ranges, or for test floor and ceiling bounds, and apply them uniformly to existing test results.

Option 1. Examine test publisher documents. Many test publishers now recommend functional testing for students and specify in their publications the reliable ranges for tests. The recommended range typically indicates a lower and upper bound for scores. A useful criterion metric is percent of raw score (number correct) since it is independent of the test level and time of testing. One publisher suggests, for example, that students should be able to answer between 40 percent and 75 percent of the items on a test if the level is to be considered appropriate.

Some publishers also provide standard errors of measurement distributions. These distributions are useful in that they provide a standard error of measurement (SEM) estimate for the amount of error associated with each raw score on a particular test. Scores below the test floor or above the test ceiling have markedly higher SEMs than score in the reliable mid-range of the test.

Option 2. Examine the raw score distribution of existing test results. The skewness of the distribution indicates the extent to which a test was appropriate for the students tested. If the mean and median differ by less than one-third of the standard deviation of scores, the test was probably appropriate for the group tested.

On the other hand, a positively skewed distribution suggests that the test was too difficult and that the test floor was encountered by a signficant number of students. Pretest floor effects reduce the potential for measuring the growth of low scoring students.

A negatively skewed distribution, however, suggest that the test was too easy and that the test ceiling was encountered. It is likely that high ability students would have obtained higher scale

score on a more difficult test. Post-test ceiling effects reduce the potential for measuring the growth of high achieving students.

Caution. Low scores (e.g., chance scores or scores at or below the test floor) may be caused by many factors and do not in and of themselves establish that the test given was too difficult for the students tested. For example, low scores may result from test administration irregularities or external interferences. Or student characteristics may have affected the scores (e.g., a student may have carelessly mismarked or transcribed answers, lacked motivation, or simply have been a slow test taker). These possibilities must be assessed before drawing final conclusions about the extent of test floor problems.

But if low scores result from mismatches between the instructional content and the test content and difficult, then they clearly indicate that the in-level testing was inappropriate for the students tested.

1.2 Develop a functional level testing proposal. Based on the evidence garnered from the systematic examination of in-level test data, decide whether functional level testing should operate for (1) every grade or selected grades only (e.g., primary or elementary); (2) every student or selected students only (e.g., compensatory education participants, limited-English-proficient students); and (3) every school or selected schools.

Decision Point #2
Generating Support

Innovators or advocates of fucntional level testing need to generate support for the testing proposal from persons with the power to prevent or facilitate its implementation. To successfully mobilize support, the innovator should identify these individuals and then appropriate address the issues and concerns they raise.

2.1 Identify the perceptions and misperceptions that key people in the system hold about functional level testing. Although these perceptions are likely to vary with the position and function of the individual, the issues and concerns raised can be expected to fall into the following categories: legitimacy, consequences, and operational problems. These issues and concerns are briefly described below and suggested approaches to each are outlined in Section 2.2.

The legitimacy of out-of-level testing (i.e., administering a test other than that nominally appropriate for a grade or grade span) may be challenged by project directors, coordinators, principals, and school board members. At issue is compliance with regulations. That is, will the external funding agencies (i.e., state and federal) accept out-of-level test data to document program effectiveness?

The perceived consequences of out-of-level testing is another issue frequently raised and often reflect misperceptions. One misperception likely to surface among compensatory education principals is that students given so-called "easier" low level test have an unfair advantage over their grade peers. That is, will students earn higher

scores, thereby reducing the number who qualify for compensatory education services and the amount of special funding received by the school?

The comparability of in-level and out-of-level testing is a closely related issue sometimes raised by district personnel. That is, will students earn very different or approximately the same scores on in-level and out-of-level tests?

Operational feasibility is the most frequently raised concern of project coordinators and school site staff. That is, is functional level testing (e.g., the administration of two or more test levels in the same classroom) an unwieldy management problem?

2.2 Prepare to address the issues and concerns raised by stakeholders. Collect appropriate documents and information and devise approaches for allaying specific concerns.

The legitimacy of functional level testing can be corroborated by many sources. Several federal and state publications now affirm the relevancy and efficacy of functional level testing. See a partial list of sources in Appendix A.

The consequences of functional level testing (i.e., earning higher scores) is a pervasive concern and is based on the invalid assumption that out-of-level test score are referenced to the norming group for whom the lower level test was designed rather than to the grade or age group of the examinee. A useful technique to overcome this concern is to demonstrate how a standard scores (i.e., scale scores) are used to translate from one test level to another. Also consider conducting a field test of out-of-level testing by administering both an in-level and an out-of-level test to a sample of students. The field test will document that (1) students given the "easier" lower level test are likely to earn higher raw scores, but (2) higher raw scores are needed to attain the scale scores that translate into a given in-level percentile rank on the "more difficult" in-level test. Research literature on out-of-level testing may also by referenced as appropriate.

The issue of in-level and out-of-level test score comparability may be addressed in several ways. However, the crux of the issue is whether or not it is legitimate to compare unreliable test scores from one test level to reliable scores from another level.

## Decision Point #3
## Selecting Test Levels

Selecting appropriate test levels for students is the most complex decision for the innovator of functional level testing. It requires consideration of many factors. For example, should the district research and evaluation specialist determine the test level selection rules or should it be a cooperative effort of district and school-level personnel? How best can teachers be given a sense of ownership without compromising the quality of the data (teaching to the test)? These decisions must be made in the context of situation factors. However, innovators may wish to consider

the following suggestions.

3.1 Consider involving curriculum/instructional experts, project coordinators and teachers in devising decision rules for selecting appropriate test levels. Teacher input is particularly valuable when determining individual student functional or competency levels.

3.2 Consult the test series publications for out-of-level testing recommendations. Most recently developed series indicate that administering test two or more levels different from the in-level testing is psychometrically acceptable (see Appenix B).

3.3 Consider developing test-text linkages. The involves (1) a detailed analysis of the facts, knowledges and types of skills tapped by each test level, (2) a detailed analysis of the instructional experiences of students, and (3) a linking of test levels to instructional levels. For example, it may be found that the instructional content of Books 7 and 8 in the XYZ Reading Series is well matched to the items and difficulty level of Level 3 of the ABC Test Series. Thus, students who by the time of testing have completed Books 7 and 8 would be assigned to Level 3 of the test series, irrespective of their grade in school.

3.4 Consider multi-level test selection. That is, consider assigning student to different test levels for different content areas (e.g., reading, mathematics). Examining past test data will indicate whether student performance in reading and mathematics is sufficiently different to warrant multi-level test selection.

3.5 Consider assigning test levels based on a combination of previous test performance and teacher judgment. The district might provide each teacher with a test level recommendation for each student based on the student's previous test scores. The teacher may then change the assignment if, in the teacher's professional judgment, the level recommended is not consistent with the student's present performance.

Decision Point #4
Administering Tests

Questions about test administration procedures are likely to arise. The case study data suggest clearly that administering several test levels in a classroom is manageable (Clarke, 1983).

4.1 Consider asking teachers participating in functional level testing in other school systems to meet with staff to discuss their procedures.

4.2 Consider recommending to teachers the following: (1) shuffling students among teachers for testing; (2) using instructional aides/resource teachers/project coordinators to administer some test levels while teachers administer others; and (3) using instructional aides to supervise one group of students while teachers test the reamining students.

Decision Point #5
Scoring and Processing

Out-of-level tests may be scored and processed in a variety of ways ranging from external scoring services that do both operations to use of district computers to hand scoring and processing at the school level. Cost, the amount of school personnel, and turn-around time should be considered when deciding which procedure to employ. Regardless of the procedure adopted, the following are suggested.

5.1 Consider conducting spot checks of the scoring and processing of out-of-level and in-level tests. This should include verifying that appropriate norming populations were used to determine the derived scores for students.

5.2 Consider analyzing out-of-level test scores separately as a part of the evaluation of functional level testing, but combining in-level and out-of-level test data for reporting purposes.

Decision Point #6
On-Site Coordination

On-site coordination is essential for the operational success of a functional level testing proposal, particularly in large school systems where out-of-level testing will be employed in several schools and teaching staff may not be familiar with testing theory or interpretation.

6.1 Consider appointing or identifying one person to coordinate testing at each school site. Provide preservice and in-service training to this individual concerning: (1) the rationale for functional level testing; (2) district policies and time lines; (3) recommended or required test level selection decision rules; and (4) possible test administration alternatives.

6.2 Encourage the on-site coordinator to translate/adapt recommendations and suggestions to condition within the school.

6.3 Encourage the on-site coordinator to provide classroom support for teachers in all operational phases, including test level selection decisions and test administration.

6.4 Consider establishing a feed-back loop to the district concerning successes and problems of functional level testing. The on-site coordinator should serve as a liason between teachers and the district.

Decision Point #7
Evaluation

Functional level testing will reduce the frustration of students given in-level tests of inappropriate content and difficulty and it is

likely to reduce teacher complaints about testing. But reduction of frustration and complaints should not be the sole criterion for judging the success of functional level testing. It is equally, if not more important that the quality of the test data dreived be such that valid conclusions can be made about school, program and instructional effectiveness. Therefore, a systematic review of test score reliability must be a continuous practice.

7.1 Consider establishing a schedule for periodic review of test data. Apply the criteria developed to establish the original need for functional level testing to these new data. The reviews may reveal a need to fine tune test level selection decision rules.

7.2 Consider developing a feed-back loop between the central office and school sites. From on-site coordinators collect information about the benefits of fucntional level testing and any operational problems encountered. Provide information back to stakeholders (e.g., project directors) concerning the technical quality of the data being gathered.

## Appendix A

### References Substantiating the Legitimacy of Functional Level Testing


Functional-level testing: Selecting the right test level. Portland, OR: ESEA, Title I Evaluation, Technical Assistance Center, Northwest Regional Education Laboratory, (undated).

Functional level testing and test score utilization. Durham, NC: ESEA, Title I Evaluation, Technical Assistance Center, NTS Research Corporation, 1980.

How to choose a functional level test: A workshop. Princeton, NJ: ESEA, Title I Evaluation, Technical Assistance Center, Educational Testing Service, (undated).

Out-of-level testing: Selecting the right test level. Portland, OR: ESEA, Title I Evaluation, Technical Assistance Center, Northwest Regional Educational Laboratory, (undated).

Test information summaries for Title I evaluation. Evanston, IL: ESEA, Title I Evaluation, Technical Assistance Center, Educational Testing Service, 1982.

Test planning guide: Suggestions for avoiding testing problems. Sacramento, CA: Office of Program Evaluation and Research, Local Evaluation Assistance Unit, California State Department of Education, 1982.

Appendix B

| Test Series | Recommendations OOL |
|---|---|
| Calfornia Achievement Tests, 70 | No more than 1 level |
| California Achievement Tests, 77-78 | Two levels |
| Circus, 72-79 | No more than 1 level |
| Comprehensive Assessment Program Series, 80 | No more than 2 level |
| CTBS, 73-75 | No more than 1 level |
| CTBS, 81-82 | No more than 2 levels |
| Cooperative Primary Tests, 1965 | Possible 1 level |
| Diagnostic Mathematics Inventory, 75-80 | Advises test level selection should be guided by correspondence between objectives taught and those of the specific test levels. |
| Durrell Listening-Reading Series, 68-70 | In-level only |
| STS Educational Development Series, 76 | No more than 1 level |
| Gates-MacGinitie Reading Tests, 65-72 | Empirical out-of-level norms available from publisher. |
| Gates-MacGinitie Reading Tests, 78 | May be administered one or more levels out-of-level Supplementary out-of-level norms available. |
| Individualized Criterion Referenced Testing, 1979 | Necessary to use EPC's scoring service to obtain NR scores for other than the five specified tests at each grade level. |
| Iowa Silent Reading Tests, 73 | Not more than 1 level out-of-level |
| Iowa Tests of Basic Skills, 1971 | One or more levels out-of-level may be appropriate. |
| Iowa Tests of Basic Skills, 1978 | May be administered one or more levels out-of-level as appropriate. |
| Keymath Diagnostic Arithmetic Test, 1971-78 | Not applicable. Manual provides procedures for establishing basal and ceiling levels to determine test length. |
| Metropolitan Achievement Tests, 70 | Recommends ool for no more than one grade below actual grade level. |
| Metropolitan Achievement Tests, 78-79 | ditto |
| Metropolitan Readiness Tests, 76 | Not necessary claims publisher |
| Nelson Reading Skills Test, 77 | One or two levels below grade level. |
| Peabody Individual Achievement Test, 70 | Test items are sequenced in increasing order of difficulty and student's basal and ceiling scores are used to determin items for individual test administration. |

| | |
|---|---|
| Prescriptive Reading Inventory, 72-77 | Level selection should be guided by correspondence betweeen objectives being taught and those of the specific test levels. |
| Reading Yardsticks, 81 | Tests administered out-of-level must be hand scored since the publisher's scoring service currently does not provide for out-of-level testing. |
| Sequential Tests of Educational Progress, 1969 | More than one level out-of-level is not recommended. |
| Sequential Tests of Educational Progress, 1979 | Out-of-level norms are avaiable. More than one level out-of-level is not recommended. |
| SRA Achievement Series, 71-77 | Publisher recommends selecting test level that is appropriate in terms of content and difficulty. |
| SRA Achievement Series, 1978 | See above. General recommendations for test level selection for average and for Title I students are made. |
| Stanford Achivement Test, 1973 | More than one level out-of-level is not recommended. |
| Stanford Diagnostic Mathematics Test, 1976 | Not recommended. If out-of-level is necessary because of correspondance to content, in-level norms for the test le level should be used. |
| Stanford Diagnostic Reading Test, 1976 | Not recommended for Title I evaluations. Out-of-level norms available from publisher are not be used for Title I. |
| Stanford Early School Achievement Test, 67-70 | Possible to test one level below grade level. |
| Sthnford Test of Academic Skills, 73 | Out-of-level testing possible. |
| Test of Acadmeic Progress, 71 | May be tested one or more levels out-of-level as necessary. |
| Test of Academic Progress, 78 | See above. |
| Tests of Basic Experineces, 70-75 | Cannot be done. |
| " " " , 78 | In-level testing only. |
| Wide Range Achievement Test, 78 | no |

Taken from Test information summaries for Title I evaluation. Evanton, IL: ESEA, Title I Evaluation, Technical Assistance Center, Educational Testing Service, 1982.